

REPORT

GERMAN RESTAURANTS IN NEW YORK

BY FABIAN BUDER

REPORT:

GERMAN RESTAURANTS IN NEW YORK

FABIAN BUDER

BACKGROUND & MOTIVATION

Clearly define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.

New York is said to be one of the most exciting cities in the world. Within an area of around 300 square miles live 8.40 million people making New York also one of the most crowded places in the US. Being the dream destination for immigrants from all over the globe, New York is having an almost infinite number of restaurants serving every cuisine one can image. Even though an overall GMP of about two trillion US dollars seems like a huge potential for businesses, the market in New York is already fairly saturated and high real estate prices leave no space for the weak, at least not in the hot spots of the city where the money lies. For a group of investors I analyze the restaurant scene in the city "that never sleeps" to check the potential for a new German restaurant and get a first idea of what would be potential locations for such a restaurant where it would not hit too much direct competition.

DATA & METHODS

Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.

I will join dataset with the geo coordinates of neighborhoods and boroughs in New York with venue data from Foursquare to analyze the competition in New York:

- City data: https://geo.nyu.edu/catalog/nyu_2451_34572 Geo coordinates to locate the boroughs and neighborhoods in New York City.
- Geo coordinates: Geopy I use the geopy package to obtain the latitude and longitude values of New York City for maps.
- Venue information: Foursquare API The Foursquare API is utilized to obtain the locations and categories of food venues in New York.

I will use the above mentioned data sources for maps and exploratory data analysis to give an overview over the competition and to identify potential spots for a new German restaurant in New York.

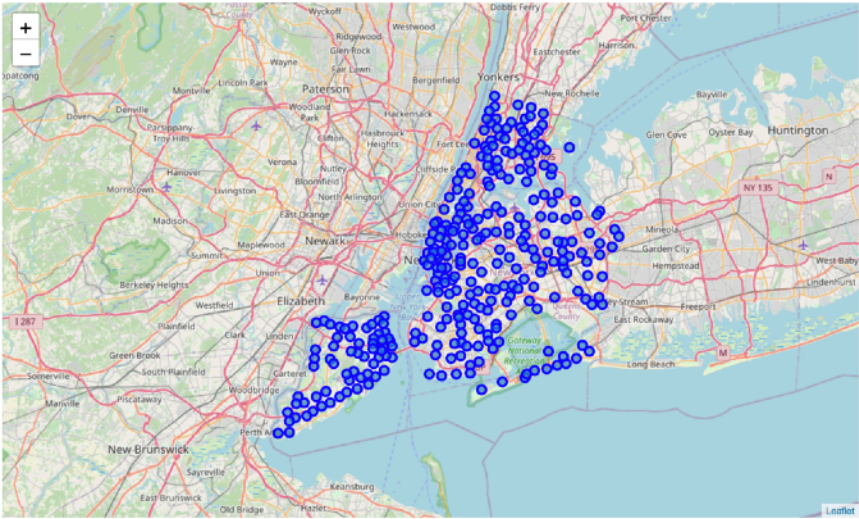
DATA ANALYSIS & RESULTS

To get started, I first need a dataset that contains the basic information about New York, the coordinates of the neighborhoods in New York that I want to check for the best place to open a new German restaurant.

New York has a total of 5 boroughs and 306 neighborhoods. In order to identify those neighborhoods in the Foursquare database later, I need a data set that contains the 5 boroughs and the neighborhoods in them together with the geodata (latitude and longitude coordinates). I download a GEOJSON file with the information from the website https://geo.nyu.edu/catalog/nyu_2451_34572.

The relevant information in the GEOJSON file is a list of the neighborhoods in New York. The information is stored in the features key. To transfer that information in a Pandas dataframe for the data analysis, I first store the information in a variable.

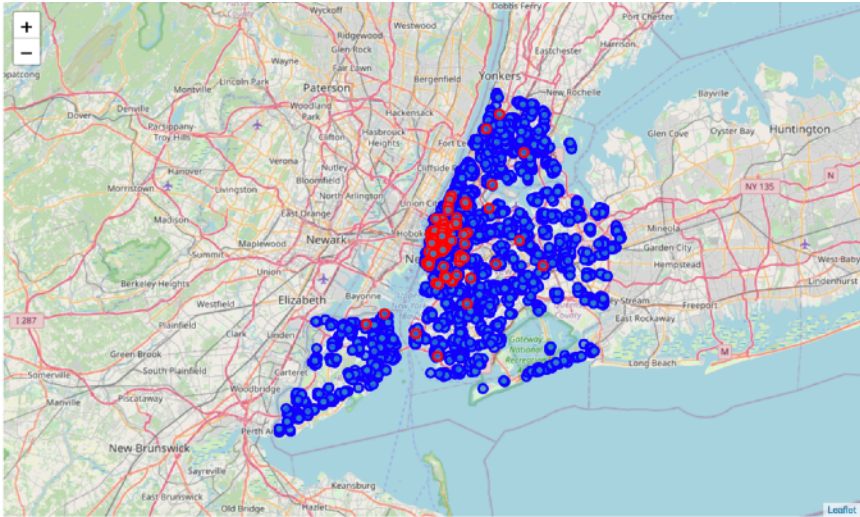
In the next step I simply transfer the variable into a pandas dataframe. Now that I have a complete list of the neighborhoods in New York, let's have a look at a map of the city and the neighborhoods in it.



Adding location data using the Foursquare Places API. To understand the competition in New York, I add the location of restaurants to the dataset.

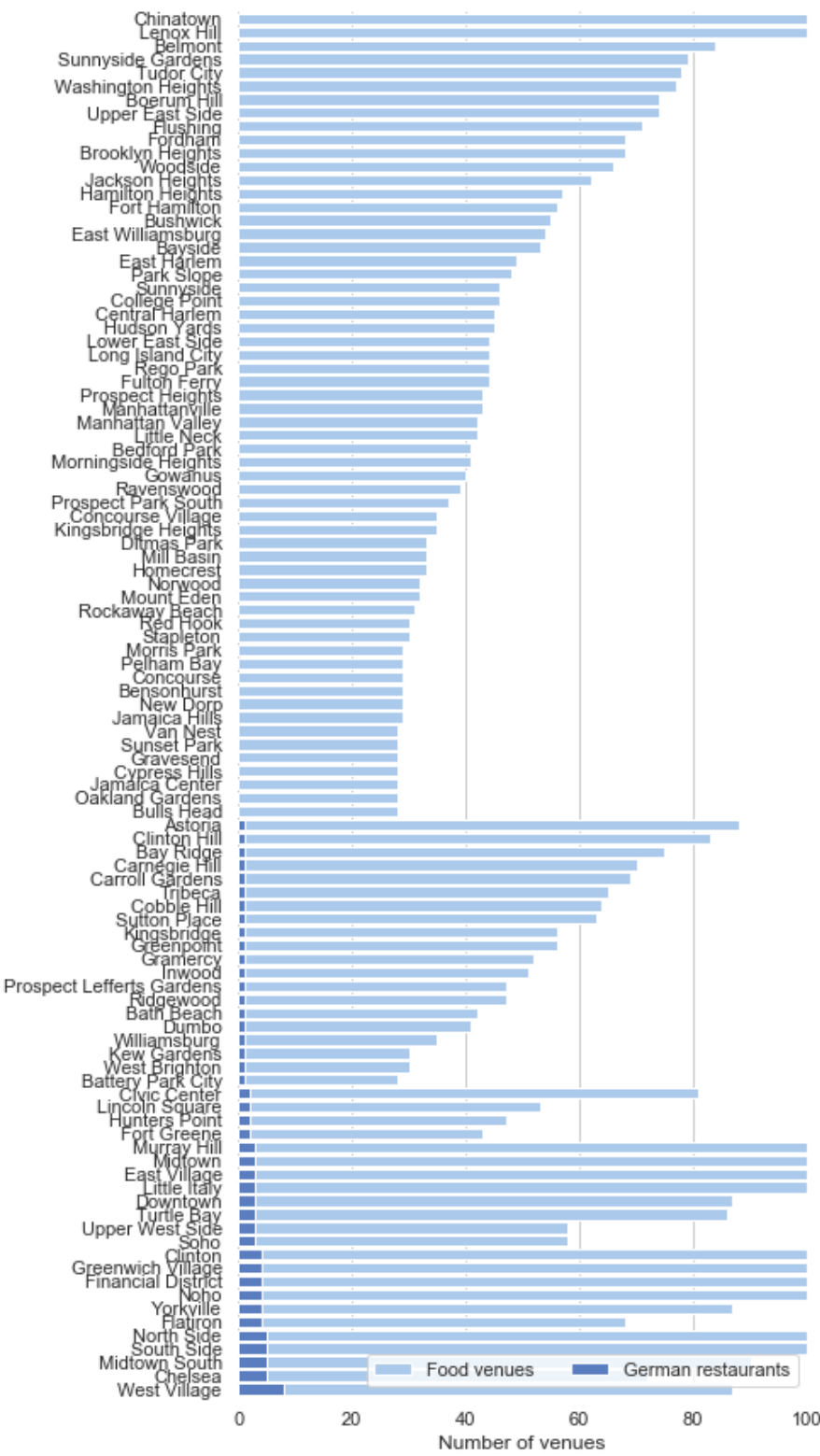
The map is good way to get an overview over neighborhoods where a certain amount of restaurants already exist, indicating that there is a customer demand for restaurants in this area and where at the same time the number of german restaurants is low or even zero.

For a better overview, I will provide a table with the necessary information on neighborhood level that allows to go into an onsite location checking phase to find the right place.



For a better overview over the neighborhoods I also prepared a bar-chart diagram on the following page.

The analysis shows that there are a lot of neighborhoods in New York that have a large number of food venues, indicating that there is a certain demand for food venues and restaurants in these areas while there is no German restaurant - at least no one that Foursquare would recommend based on the user tips.



DISCUSSION

This analysis provides a list of areas with a high density of food venues and low number of German restaurants as a starting point for a further, more customer oriented analysis of the areas.

While this is this list and the maps are good starting point to look for areas with high potential for a German restaurant there are certain limitations due to the used dataset. The approach of looking for venues in a certain range around a given geo may lead to incomplete lists of venues. Focussing only on recommended venues also may shorten the list. Nonetheless - if we aim for a restaurant with a great customer experience, low-rated restaurants nearby may be no real competition.

Further research is needed to analyse the buying power and cultural orientation in the neighborhoods to find the perfect place for a new restaurant. Some of the neighborhoods that have no German restaurant may have none for a reason - from a naive perspective for example China Town may not be the first place to think about opening a German restaurant.

CONCLUSION

This test case led to some nice maps and already produced some good results. Real world use cases would probably work with clearer expectations towards the right spot for a new branch and allow to search more goal oriented. Also some of the limitations of the dataset may be corrected by adding additional resources into programming.

Thank you for reviewing my work. I enjoyed the course but I also clearly see that way more training is necessary to get used to the Python language and work way faster and more efficient.